

UNITED STATES PATENT APPLICATION FOR

NROM CELL WITH N-LESS CHANNEL

Inventors:

MASAAKI HIGASHITANI

MARK RANDOLPH

10000112-0007002

# NROM CELL WITH N-LESS CHANNEL

## BACKGROUND

The present invention relates generally to methods of fabrication of nitride read only memory (NROM) cells and arrays.

FIG. 1 illustrates a typical NROM cell. This cell includes a substrate 10 in which are implanted a bit line source 12 and drain 14 and on which lies an oxide-nitride-oxide (ONO) structure 16 having a layer of nitride 17 sandwiched between a lower oxide layer 18 and an upper oxide layer 20. On the ONO structure 16 lies a polysilicon connecting block 26, which functions as a gate conductor. On the source 12 and drain 14 are isolated oxide areas 62, which function as insulators. Between the source 12 and drain 14 is a channel 15 under the ONO structure 16.

The nitride section 17 provides the charge retention mechanism for programming the memory cell. Specifically, when programming voltages are provided to the source 12, drain 14 and gate conductor 26, electrons flow towards the drain 14. According to the hot electron injection phenomenon, some hot electrons penetrate through the lower section of silicon oxide 18, especially if the section 18 is thin, and are then collected in the nitride section 17. As is known in the art, the nitride section 17 retains the received charge, labeled 24, in a concentrated area adjacent the drain 14. The concentrated charge 24 significantly raises the threshold of the portion of the channel of the memory cell under the charge 24 to be higher than the threshold of the remaining portion of the channel 15.

When concentrated charge 24 is present (i.e. the cell is programmed), the raised threshold of the cell does not permit the cell to be placed into a conductive state during reading of the cell. If concentrated charge 24 is not present, the read voltage on gate conductor 22 can overcome the much lower threshold and accordingly, channel 15 becomes inverted and hence, conductive. It is noted that the threshold voltage  $V_{th}$  of NROM cells is generally very sensitive to the voltages  $V_{drain}$  and  $V_{gate}$  provided on the drain 14 and on the gate conductor 26, respectively.

FIGS. 2-5 illustrate a conventional approach to forming the NROM cell of FIG. 1. The process begins by thermally growing a lower layer of silicon oxide 18 on a P-type silicon substrate 10. Then, a silicon nitride layer 17 is deposited by LPCVD and the wafer is steam-sealed to form the upper silicon oxide layer 20. Next, a doped layer of polysilicon is

deposited over the wafer to a depth of about 4,000 angstroms. The polysilicon layer is patterned to form polysilicon connecting blocks 26, with openings 30 in between.

A heavy dosage of arsenic 36 is then implanted into the wafer to form heavily doped, N-type, bit line areas 38. The polysilicon connecting blocks 26 are used for connecting to  
5 conductive word lines which are common to other similar memory cells in the rows. Also, the polysilicon connecting blocks 26 function as masks to prevent the arsenic implant from reaching the underlying channel areas 15.

The wafer is then subjected to an etch for removing portions of the oxide-nitride-oxide layers 18, 17, and 20 which are not covered by the polysilicon connecting blocks 26.  
10 The remaining oxide-nitride-oxide layers 18, 17, and 20 are self-aligned with respect to the side edges of the overlying polysilicon connecting blocks 26, as illustrated in FIG. 2.

The process continues with the deposition of a layer of conformal silicon oxide 52, such as tetra-ethoxysilane (TEOS), over the surface of the wafer to a depth of about 5,000 angstroms. A layer of photoresist 54 is then spun over the surface of the wafer for filling the  
15 contoured surface of the conformal silicon oxide layer 52, and for providing a flat top surface. The wafer may be heated to a temperature effective to anneal the implant and activate the impurities to form diffused bit line regions 56, shown in FIG. 3. The diffused bit line regions 56 are elongate and are associated with other cells in the columns of the memory array.

20 An etching process is then employed for etching the photoresist 54 and the conformal silicon oxide layer 52 at the same rate. When such an etch is conducted, the removal of material proceeds uniformly downwardly, until the polysilicon connecting blocks 26 are reached. The top surface of the polysilicon connecting blocks 26 are planarized with respect to the top surface of the conformal silicon oxide layer 52, forming isolated oxide areas 62.  
25 The result is a planarized surface of the memory array, illustrated in FIG. 4.

Finally, a second layer of doped polysilicon 68 is deposited over the surface of the wafer, as shown in FIG. 5, and patterned to define a word line extending in common with a number of other memory cells of the row. Importantly, the polysilicon word line 68 is in electrical contact with the first polysilicon layer forming the conductive connecting blocks  
30 26. Thus, when an address signal is applied to the word line 68, such signal is applied simultaneously to the polysilicon connecting blocks 26 which function as gate conductors. All memory cells connected to the word line 68 which are not programmed with a concentrated charge on the silicon nitride layer 17 will conduct and present a low impedance

between the associated pair of bit lines. Those memory cells in the row which are programmed so as to have a concentrated charge on the silicon nitride layer 17 will not be made conductive and thus will present a high impedance between the associated bit lines.

## 5 BRIEF SUMMARY

In one aspect, the present invention is an electrically programmable read-only memory device including an array of single transistor memory cells where each of said cells is read, programmed, and erased through a pair of associated bit lines and a single associated word line. The memory device is formed on a substrate. Each memory cell comprises a pair  
10 of bit lines extending in a first direction across the substrate, a pair of bit line dielectrics overlaying and covering the pair of bit lines, a charge-trapping layer formed over the channel region between the pair of bit lines, and a conductive connecting block formed on the charge-trapping layer. The charge-trapping layer comprises two oxide-nitride-oxide  
15 (ONO) structures separated by a gate oxide layer, where each ONO structures comprises a layer of nitride sandwiched between a bottom oxide layer and a top oxide layer. A plurality of straight, parallel-edge-word lines extend across the substrate in a second direction and cross over the bit lines and channel regions. Each word line comprises a conductive material and is separated from the substrate by the conductive connecting blocks and bit line dielectrics.

20 In a second aspect, the present invention concerns a first method of fabricating a nitride read only memory (NROM) chip. The method comprises creating an oxide-conductive layer on a substrate, where the oxide-conductive layer is formed of a first conductive layer on top of a thick oxide layer; laying down a bit line mask of photoresist generally in columns at least within a memory portion of the chip; removing at least a  
25 portion of the oxide-conductive layer wherever the photoresist is not present to form oxide-conductive-layer columns; implanting bit lines wherever the photoresist is not present and generally in columns; removing the photoresist; removing a portion of the remaining thick oxide layer from the oxide-conductive-layer columns to produce a recess; growing a thin oxide layer over the memory portion of the chip; depositing a nitride layer over the thin  
30 oxide layer to a thickness sufficient to fill the recess; removing the nitride layer except in the region of the recess; forming bit line dielectrics on top of the bit lines; and forming rows of a second conductive layer perpendicular to and on top of the bit line dielectrics and oxide-conductive-layer columns.

10000112-000702

In a third aspect, the present invention concerns a second method of fabricating a nitride read only memory (NROM) chip. The method comprises creating an oxide-conductive layer on a substrate, where the oxide-conductive layer is formed of a first conductive layer on top of a thick oxide layer; laying down a bit line mask of photoresist, the bit line mask formed generally in columns at least within a memory portion of the chip; removing at least a portion of the oxide-conductive layer wherever the photoresist is not present to form a plurality of oxide-conductive-layer columns; implanting bit lines wherever the photoresist is not present and generally in columns; removing the photoresist; forming an oxide-nitride-oxide (ONO) layer within the memory portion of the chip, where the ONO layer comprises a bottom oxide layer, a nitride layer, and a top oxide layer; removing the ONO layer except from the regions adjacent to the oxide-conductive-layer columns to form pairs of ONO structures separated by oxide-conductive-layer columns; forming bit line dielectrics on top of the bit lines and ONO structures; and forming rows of a second conductive layer perpendicular to and on top of the bit line dielectrics and oxide-conductive-layer columns.

#### DEFINITIONS:

The term "substrate" refers to any semiconductor material conventionally known to those of ordinary skill in the art. Examples include silicon, gallium arsenide, germanium, gallium nitride, aluminum phosphide, and alloys such as  $\text{Si}_{1-x}\text{Ge}_x$  and  $\text{Al}_x\text{Ga}_{1-x}\text{As}$ , where  $0 \leq x \leq 1$ . Many others are known, such as those listed in Semiconductor Device Fundamentals, on page 4, Table 1.1 (Robert F. Pierret, Addison-Wesley, 1996). Preferably, the semiconductor substrate is silicon, which may be doped or undoped.

The term "planarizing" means to remove material, preferably in the vertical direction, enhancing the flatness of the structure being planarized.

The term "adjacent" means that there are no functional structures between the specified structures. A functional structure is a structure intentionally placed on a semiconductor device that affects the function of the device. For example, a functional structure may mean a structure having a predetermined set of dimensional and/or compositional parameter values that has an electrical, mechanical and/or optical function (e.g., conductive, insulative, masking, photolithographic, or antireflective).

The term "oxide" refers to a metal oxide conventionally used to isolate electrically active structures in an integrated circuit from each other, typically an oxide of silicon and/or aluminum (e.g.,  $\text{SiO}_2$  or  $\text{Al}_2\text{O}_3$ , which may be conventionally doped with fluorine, boron,

phosphorus or a mixture thereof; preferably  $\text{SiO}_2$  or  $\text{SiO}_2$  conventionally doped with 1-12 wt% of phosphorous and 0-8 wt% of boron).

The term "dielectric layer" refers to any dielectric material conventionally known to those of ordinary skill in the art. Examples include conventional oxides, nitrides, oxynitrides, and other dielectrics, such as borophosphosilicate glass (BPSG), borosilicate glass (BSG), phosphosilicate glass, spin-on glass (SOG), silicon oxide, P-doped silicon oxide (P-glass), and silicon nitride, for example  $\text{SiO}_2$ ,  $\text{Si}_3\text{N}_4$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{SiO}_x\text{N}_y$ , etc.

## BRIEF DESCRIPTION OF SEVERAL VIEWS OF THE DRAWINGS

Various other objects, features and attendant advantages of the present invention will be more fully appreciated as the same becomes better understood from the following detailed description when considered in connection with the accompanying drawings in which like reference characters designate like or corresponding parts throughout the several views and wherein:

FIG. 1 is a schematic illustration of a prior art NROM memory cell;

FIGS. 2-5 illustrate sectional views of the NROM cell of FIG. 1, taken during various process steps of the fabrication thereof;

FIG. 6 is a schematic illustration of the NROM memory cell according to a first embodiment of the present invention;

FIG. 7-13 illustrate sectional views of the NROM cell of FIG. 6, taken during various process steps of the fabrication thereof;

FIG. 14 is a schematic illustration of the NROM memory cell according to a second embodiment of the present invention;

FIG. 15-20 illustrate sectional views of the NROM cell of FIG. 14, taken during various process steps of the fabrication thereof.

## DETAILED DESCRIPTION OF THE PRESENTLY PREFERRED EMBODIMENTS

FIG. 6 illustrates an NROM cell according to a first embodiment of the present invention. This cell includes a substrate 110 in which are implanted source/drain bit lines 148 and on top of which lies a charge trapping layer, comprising a pair of oxide-nitride-oxide (ONO) structures 116 separated by a gate oxide layer 120. Preferably, the first ONO structure 116 is on a first portion of the substrate 110, the second ONO structure 116 is on a

second portion of the substrate 110, wherein the first portion does not completely overlap the second portion, as illustrated in FIG. 6. The first portion may overlap a portion of the second portion, so long as the first portion does not overlap the entire second portion. More preferably, the gate oxide 120 is on a third portion of the substrate, wherein the third portion does not completely overlap either the first or second portion. The first ONO structure 116 and the second ONO structure 116 do not form a stacked structure, that is, a structure wherein the first ONO layer 116 is on the second ONO structure 116. But rather, the first ONO structure 116 is laterally aligned with the second ONO structure 116, wherein the gate oxide layer 120 is located in between the first and second ONO structures 116, as illustrated in FIG. 6.

Gate oxide layer 120 includes any dielectric material conventionally known to those of ordinary skill in the art. Examples of dielectric material include conventional oxides, nitrides, oxynitrides, and other dielectrics, such as borophosphosilicate glass (BPSG), borosilicate glass (BSG), phosphosilicate glass, spin-on glass (SOG), silicon oxide, P-doped silicon oxide (P-glass), and silicon nitride, for example  $\text{SiO}_2$ ,  $\text{Si}_3\text{N}_4$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{SiO}_x\text{N}_y$ , etc. Each ONO structure comprises a nitride layer 117 sandwiched in a thin oxide layer 118. The charge-trapping layer is located above the channel 115, defined as that portion of the surface of the substrate 110 lying between a pair of bit lines 148. On top of the charge trapping layer lies a conductive connecting block 126, which functions as a gate conductor. On top of the bit lines 148 lie bit line dielectrics 150, which function as insulators. In this embodiment, the conductive connecting block 126 overlays and covers both the gate oxide layer 120 and the ONO structures 116.

Reference is now made to FIGS. 7-13, which illustrate the NROM fabrication method according to a first embodiment of the present invention. The process begins by growing a thick oxide layer 121 on a substrate 110, preferably to a thickness of 150-200 Å, and preferably in a low temperature oxidation operation. A preferable oxidation temperature is about 800 degrees C, but it can vary between 750 and 1000 degrees C. The substrate 110 is preferably a P-type silicon substrate of a <100> crystal lattice orientation.

Then, a first conducting layer 140 is deposited over the wafer to a depth of 3000-5000 Å, resulting in the structure in Figure 7. The first conducting layer is preferably polycrystalline silicon, more preferably doped polysilicon.

The next step involves depositing a photoresist layer and patterning it into columns 138 to form a bit line mask 136, whose layout within the memory array portion of the chip is

shown from above in FIG. 8A. The photoresist columns 138 define the areas where the bit lines are not to be implanted (i.e. the locations of the channels 115 (FIG. 6)). Prior to implanting the bit lines, the thick oxide and first conducting layers 121 and 140, respectively, are etched away from the areas between photoresist columns 138 to form conductive connecting blocks 126 superposed on oxide blocks 144. The conductive connecting blocks 126 and oxide blocks 144 are self-aligned with respect to the side edges of the overlying photoresist columns 138, as illustrated in FIG. 8B.

After the etch operation, bit lines 148 are implanted (FIG. 8C) in the areas between the photoresist columns 138. A typical implant might be  $2-4 \times 10^{15}$  /cm<sup>2</sup> of Arsenic at 50 Kev. It will be appreciated that this is a self-aligned implant in which the bit lines 148 are self-aligned to the photoresist columns 138. In addition, the bit line regions 148 are elongate and are associated with other cells in the columns of the memory array. Next, the first photoresist layer 138 is stripped and the surface is cleaned.

An isotropic HF etch is then performed to remove a portion of the oxide block 144 and produce a recess 119 and a gate oxide layer 120 over a channel 115, as illustrated in FIG. 9. The amount of oxide removed is typically 0.02-0.05  $\mu$ , or approximately  $\frac{1}{4}$  to  $\frac{1}{2}$  of the width of the conductive connecting block 126. If too much oxide is removed, the polysilicon connecting blocks 126 becomes unstable and can topple over.

A thin oxide layer 118 is then grown over the wafer, typically to a thickness of 40-70 Å in a low temperature oxidation operation. A preferable oxidation temperature is about 800° C, but it can vary between 750-1000° C. A silicon nitride layer 117 is next deposited over the thin oxide layer 118 in a low temperature chemical vapor deposition (LPCVD) operation to a thickness of 50-70 Å. Importantly, the nitride layer 117 is grown to a thickness sufficient to fill the recess 119 under the conductive connecting blocks 146 resulting from the prior isotropic HF oxide etch. The LPCVD operation is typically conducted at a temperature of 650-750° C and a pressure of 0.05-0.3 Torr using a gas mixture of NH<sub>3</sub> and SiCl<sub>2</sub>H<sub>2</sub>. Preferably, the temperature is 725° C and the pressure is 0.2 Torr. The resulting structure is depicted in FIG. 10.

Afterwards, an anisotropic etch is performed to remove the nitride layer 117 from the surface of the chip, except from the recessed areas 119 (shown in FIG. 9). The thin oxide layer 118 is left essentially unperturbed by this operation. As shown in FIG. 11, this etch generates a charge-trapping layer located above a channel 115 and under a conductive connecting block 126, wherein the charge trapping layer comprises a pair of oxide-nitride-



oxide (ONO) columns 116 separated by a gate oxide 120. Each ONO column comprises a nitride layer 117 sandwiched in a thin oxide layer 118.

This construction of the charge-trapping layer minimizes the cross-talk between source/drain bit line 148 pairs often associated with previous structures. As a result, the reliability and dynamic range (i.e. voltage required for a read/write operation) of the resulting device are significantly enhanced. Concurrently, the programming and erase time are advantageously decreased.

The process continues with the deposition of a dielectric layer, preferably conformal silicon oxide, over the surface of the chip to a depth of 4000-6000 Å. Preferably, the deposition uses a high-density plasma or a conventional CVD method, for example, the thermal decomposition of tetraethyl orthosilicate (TEOS). Chemical mechanical polishing (CMP) is then employed to uniformly remove material in a downward direction until the conductive connecting blocks 126 are reached. The result is a planarized surface of the memory array, with bit line dielectrics 150 formed over the bit lines 148, illustrated in FIG 12.

With reference now to FIG. 13A, a second conducting layer 152, preferably polycrystalline silicon, more preferably doped polysilicon, is deposited over the surface of the chip. Then, the conductive connecting blocks 126 and second conducting 152 layer are concurrently patterned to define a word line 154 (FIG. 13B viewed from above) extending in common with a number of other memory cells of the row. Importantly, the word line 154 is in electrical contact with the first conductive layer 140 forming the conductive connecting blocks 126. Thus, when an address signal is applied to the word line 154, such signal is applied simultaneously to the conductive connecting blocks 126, which function as gate conductors. All memory cells connected to the word line 154 which are not programmed with a concentrated charge on the silicon nitride layer 117 will conduct and present a low impedance between the associated pair of bit lines. Those memory cells in the row which are programmed so as to have a concentrated charge on the silicon nitride layer 117 will not be made conductive and thus will present a high impedance between the associated bit lines.

FIG. 14 illustrates an NROM cell according to a second embodiment of the present invention. This cell includes a substrate 210 in which are implanted source/drain bit lines 248 and on top of which lies a charge trapping layer, comprising a pair of oxide-nitride-oxide (ONO) structures 216 separated by a gate oxide 220. Preferably, the first ONO structure 216 is on a first portion of the substrate 210, the second ONO structure 216 is on a

second portion of the substrate 210, wherein the first portion does not completely overlap the second portion, as illustrated in FIG. 14. The first portion may overlap a portion of the second portion, so long as the first portion does not overlap the entire second portion. More preferably, the gate oxide 220 is on a third portion of the substrate, wherein the third portion does not completely overlap either the first or second portion. The first ONO structure 216 and the second ONO structure 216 do not form a stacked structure, that is, a structure wherein the first ONO layer 216 is on the second ONO structure 216. But rather, the first ONO structure 216 is laterally aligned with the second ONO structure 216, wherein the gate oxide layer 220 is located in between the first and second ONO structures 216, as illustrated in FIG. 14.

Gate oxide 220 includes any dielectric material conventionally known to those of ordinary skill in the art. Examples of dielectric material include conventional oxides, nitrides, oxynitrides, and other dielectrics, such as borophosphosilicate glass (BPSG), borosilicate glass (BSG), phosphosilicate glass, spin-on glass (SOG), silicon oxide, P-doped silicon oxide (P-glass), and silicon nitride, for example  $\text{SiO}_2$ ,  $\text{Si}_3\text{N}_4$ ,  $\text{Al}_2\text{O}_3$ ,  $\text{SiO}_x\text{N}_y$ , etc. Each ONO structure comprises a nitride layer 217 sandwiched between a bottom oxide layer and a top oxide layer. The charge trapping layer is located above the channel 215, as well as over the ends of the source/drain regions 248. On top of the charge trapping layer lies a conductive connecting block 226, which functions as a gate conductor. On top of the bit lines 248 and ONO structures 216 lie bit line dielectrics 250, which function as insulators. In this embodiment, the conductive connecting block 226 overlays and covers only the gate oxide layer 220.

Reference is now made to FIGS. 15-20, which illustrate the NROM fabrication method according to a second embodiment of the present invention. The process begins by growing a thick oxide layer 221 on a substrate 210, preferably to a thickness of 150-200 Å, in a low temperature oxidation operation. A preferred oxidation temperature is about 800 degrees C, but it can vary between 750 and 1000 degrees C. The substrate 210 is preferably a P-type silicon substrate of a <100> crystal lattice orientation.

Then, a first conducting layer 240 is deposited over the wafer to a depth of 3000-5000 Å, resulting in the structure in Figure 15. The first conducting layer is preferably polycrystalline silicon, more preferably doped polysilicon.

The next step involves depositing a photoresist layer and patterning it into columns 238 to form a bit line mask 236, whose layout within the memory array portion of the chip is

shown from above in FIG. 16A. Prior to implanting the bit lines, the thick oxide and first conducting layers 221 and 240, respectively, are etched away from the areas between photoresist columns 238 to form conductive connecting blocks 226 superposed on gate oxide layers 220. The conductive connecting blocks 226 and gate oxide layers 220 are self-aligned with respect to the side edges of the overlying photoresist columns 238, as illustrated in FIG. 16B.

After the etch operation, bit lines 248 are implanted (FIG. 16C) in the areas between the photoresist columns 138. A typical implant might be  $2-4 \times 10^{15}$  /cm<sup>2</sup> of Arsenic at 50 Kev. It will be appreciated that this is a self-aligned implant in which the bit lines 248 are self-aligned to the photoresist columns 238. In addition, the bit line regions 248 are elongate and are associated with other cells in the columns of the memory array. Next, the first photoresist layer 238 is stripped and the surface is cleaned.

A lower oxide layer 218 is then grown over the chip, preferably to a thickness of 10-50 Å in a low temperature oxidation operation. A preferred oxidation temperature is about 800° C, but it can vary between 750-1000° C. A silicon nitride layer 217 is next deposited over the thin oxide layer 218 in a low temperature chemical vapor deposition (LPCVD) operation to a thickness of 50-70 Å. The LPCVD operation is preferably conducted at a temperature of 650-750° C and a pressure of 0.05-0.3 Torr using a gas mixture of NH<sub>3</sub> and SiCl<sub>2</sub>H<sub>2</sub>. Preferably, the temperature is 725° C and the pressure is 0.2 Torr. Then, an upper oxide layer 220 is grown over the nitride layer 217 to a depth of 50-150 Å. The resulting structure is depicted in FIG. 17.

Afterwards, an isotropic etch is performed to remove the upper oxide, nitride, and lower oxide layers 220, 217, and 218, respectively, from the surface of the chip except for the regions adjacent to the conductive connecting blocks 226. As shown in FIG. 18, this etch generates a charge trapping layer located above a channel 115. The charge-trapping layer comprises a pair of oxide-nitride-oxide (ONO) columns 216 separated by a gate oxide 220, wherein each ONO column comprises a nitride layer 217 sandwiched between a lower oxide layer 218 and an upper oxide layer 220. The conductive connecting block 226 overlays and covers only the gate oxide 220.

The process continues with the deposition of a dielectric layer, preferably conformal silicon oxide, over the surface of the chip to a depth of 4000-6000 Å. Preferably, the deposition uses a high-density plasma or a conventional CVD method, for example, the thermal decomposition of tetraethyl orthosilicate (TEOS). Chemical mechanical polishing

(CMP) is then employed to uniformly remove material in a downward direction until the conductive connecting blocks 226 are reached. The result is a planarized surface of the memory array, with bit line dielectrics 250 formed over the bit lines 248, illustrated in FIG 19.

5 With reference now to FIG. 20A, a second conducting layer 252, preferably polycrystalline silicon, more preferably doped polysilicon, is deposited over the surface of the chip. Then, the conductive connecting blocks 226 and second conducting layer 252 are concurrently patterned to define a word line 254 (FIG. 20B) extending in common with a number of other memory cells of the row. Importantly, the word line 254 is in electrical  
10 contact with the first conductive layer 240 forming the conductive connecting blocks 226. Thus, when an address signal is applied to the word line 254, such signal is applied simultaneously to the conductive connecting blocks 226, which function as gate conductors. All memory cells connected to the word line 254 which are not programmed with a  
concentrated charge on the silicon nitride layer 217 will conduct and present a low

15 impedance between the associated pair of bit lines. Those memory cells in the row which are programmed so as to have a concentrated charge on the silicon nitride layer 217 will not be made conductive and thus will present a high impedance between the associated bit lines.

It is noted that the present invention covers the fabrication of the entire chip which includes the NROM memory array portion and the complementary metal oxide  
20 semiconductor (CMOS) periphery devices.

The individual processing steps for use in the present invention are well known to those of ordinary skill in the art, and are also described in Encyclopedia of Chemical Technology, Kirk-Othmer, Volume 14, pp. 677-709 (1995); Semiconductor Device Fundamentals, Robert F. Pierret, Addison-Wesley, 1996; and Microchip Fabrication 3<sup>rd</sup>  
25 edition, Peter Van Zant, McGraw-Hill, 1997.

The dielectric layer may be deposited by conventional methods known to those of ordinary skill in the art, such as by spin-on methods, sintering (which may further include sol-gel oxide formation), chemical vapor deposition, etc. A glass layer deposited by a chemical vapor deposition technique may be subject to a glass reflow step (e.g., by heating)  
30 to smooth, densify and further improve the contact between the protection layer and the substrate.

Etching of deposited films may be conducted by conventional methods known to those of ordinary skill in the art. The specific etching methods and materials depend on the

material being removed, the resist material and the compatibility of the etching material with the existing structure. Selection of suitable etching materials, resist materials and etching conditions is within the level of ordinary skill in the art.

5 The semiconductor structures of the present invention may be incorporated into a semiconductor device such as an integrated circuit, for example a memory cell such as an SRAM, a DRAM, an EPROM, an EEPROM etc.; a programmable logic device; a data communications device; a clock generation device; a nonvolatile memory device, etc. Furthermore, any of the semiconductor devices may be incorporated into an electronic device (e.g. a computer, an airplane, a mobile telephone, or an automobile).

10 Thus, there has been disclosed in accordance with the invention, an apparatus and method for creating a semiconductor structure that fully provides the advantages set forth above. Although the invention has been described and illustrated with reference to specific illustrative embodiments thereof, it is not intended that the invention be limited to those illustrative embodiments. Those skilled in the art will recognize that variations and  
15 modifications can be made without departing from the spirit of the invention. It is therefore intended to include within the invention all such variations and modifications that fall within the scope of the appended claims and equivalents thereof.